

Models for unitary black hole disintegration

STEVEN B. GIDDINGS*

*Department of Physics
University of California
Santa Barbara, CA 93106*

Abstract

Simple models for unitary black hole evolution are given in an effective Hilbert-space description, parameterizing a possible minimal relaxation of locality, with respect to semi-classical black hole geometry.

* Email address: giddings@physics.ucsb.edu

1. Introduction

The unitarity crisis or black hole (BH) information problem challenges the pillars of modern physics: quantum mechanics (QM), Lorentz/diffeomorphism invariance, and locality. In short, Hawking’s argument for BH evaporation[1] yields information loss[2], and it has been argued that there is no consistent scenario with these pillars intact.¹

Locality, in particular, seems a weak link in quantum gravity, and there have been suggestions to modify it[8-13] in some way. However, such modification should be subtle as locality is a basic principle of local quantum field theory (LQFT), which describes observed phenomena extremely well. Generally, we observe that quantum information is well localized in spacetime, and moreover generic nonlocality leads to causality paradoxes.

One thus seeks a consistent dynamics describing deeper relations between quantum information and spacetime and its symmetries, where locality may be approximate but not exact, yet consistent unitary evolution is intact.

Assuming QM, a useful tool is an effective quantum information-theoretic parameterization, with a simple Hilbert-space description of the dynamics. Such an approach has been used *e.g.* in [14-17,7,18]. Approximate locality suggests we decompose the overall Hilbert space into a product of Hilbert spaces $\hat{\mathcal{H}}$ and \mathcal{H} corresponding to states inside and outside a BH, with evolution providing specific “small” couplings between these.

Such a framework is general enough to describe local evolution, but also nonlocal modifications. Our approach will seek a conservative alternative[13] to complementarity/holography[9,10], staying as close as possible to LQFT, with minimal, controlled allowance for nonlocality, as needed for unitary evolution. In particular, in the evolution of [1], $\hat{\mathcal{H}}$ is effectively infinite dimensional, but unitarity and other indicators suggest an effective number of BH degrees of freedom $N(M) = \log(\dim \hat{\mathcal{H}})$ at BH mass M that is finite, *e.g.* given by the Bekenstein-Hawking[19,1] entropy,

$$N(M) = S_{BH}(M) . \tag{1.1}$$

A problem is how to describe physical unitary evolution incorporating such finite information content for a BH.

We will begin by reviewing the familiar scenario of Hawking evaporation[1], to make contact with such an effective Hilbert-space description. Then, we will describe models for unitary but nonlocal evolution. Such models at the least may guide understanding of the constraints on unitary scenarios, but might also guide deeper understanding of the principles of quantum gravity.

¹ For reviews see [3-7].

2. Hawking evaporation

We take rotation to be an inessential complication and consider a Schwarzschild metric,

$$ds^2 = -f(r)dt^2 + \frac{dr^2}{f(r)} + r^2 d\Omega^2 . \quad (2.1)$$

This form is valid in general spacetime dimension D , although one may particularly focus on $D = 4$ where $f(r) = 1 - 2M/r$. Hawking radiation is conveniently analyzed by introducing tortoise coordinates, via $dr^*/dr = 1/f(r)$. Near infinity, $r^* \approx r$, but as the horizon $r = R$ is approached, $r^* \rightarrow -\infty$. Consider a free field Φ , with general spin. Solutions of the corresponding wave equation may be written as a sum of terms

$$\Phi_J = \frac{u_J(r, t)}{r^{D/2-1}} Y_J(\Omega) , \quad (2.2)$$

where J labels angular momentum and Y_J is an appropriate spherical harmonic. In terms of (r^*, t) , u_J satisfies a two-dimensional wave equation with effective potential that vanishes at $r^* = \pm\infty$, and has a maximum $\sim J^2/R^2$. Further details appear in, *e.g.*, [4].

The quantum field outside the BH may be expanded in modes $\Phi_{J,\omega}$, of energy ω , and creation/annihilation operators, as $\Phi = \sum_J \int d\omega (\Phi_{J,\omega} a_{J,\omega} + h.c.)$. As reviewed in [4], the Hawking state is found by determining the transformation between the coordinate $x^- = t - r^*$, in which the “out” vacuum is naturally defined, and appropriate coordinates for defining an “in” vacuum. One may also choose corresponding modes [20,4] $\hat{\Phi}_{J,\omega}$ inside the horizon, with mode operators $\hat{a}_{J,\omega}$. Then, the “in” vacuum evolves to the Hawking state, which takes the form

$$|\psi\rangle = c \sum_{\{n_{J,\omega}\}} e^{-H/2T} |\{\hat{n}_{J,\omega}\}\rangle |\{n_{J,\omega}\}\rangle . \quad (2.3)$$

Here $n_{J,\omega}$ are occupation numbers, $T = (D-3)/4\pi R$ is the temperature,

$$H = \sum_J \int d\omega \omega n_{J,\omega} \quad (2.4)$$

is the hamiltonian in the Fock basis we have chosen, and c is a constant. Tracing out the internal states gives a thermal density matrix.

In order to match to an effective Hilbert space description, it is useful to choose a wavepacket basis. For example, a simple set of complete, orthonormal states investigated in [1,20] is, with integer k and n ,

$$u_{J,k,n} = \epsilon^{-1/2} \int_{k\epsilon}^{(k+1)\epsilon} d\omega e^{2\pi i \omega n / \epsilon} u_{J,\omega} . \quad (2.5)$$

These have frequency $\omega \simeq k\epsilon$ and are localized about $x^- = 2\pi n/\epsilon$ with width $1/\epsilon$. The variable ϵ is an arbitrary choice; it is convenient to take $\epsilon \gtrsim 1/R$. Other, smoother, wavepacket bases may be chosen, but (2.5) is simple and intuitive. The hamiltonian H may be easily reexpressed in such a basis.

We can now describe an effective Hilbert-space model for evolution governed by H , following discussion of [14-17,7,18]. It is convenient to do so by describing the modes and their state in terms of a time slicing. One such choice is a nice slicing, as described in [21,15,7,13]; in the particular realization of [13] the slice S_t asymptotes to the constant t slice at $r = \infty$; inside the horizon and in the far past it asymptotes to $r = r_c$. In the static BH geometry slices at different t are just t -translates of this slice; with decreasing BH mass minor adjustments to such slices are needed.

On such slices, evolution of the state (2.3) in a basis such as (2.5) may be pictured as follows. First, excited quanta at $r \gg R$ simply evolve outwards. Quanta at $r < R$ evolve inwards; in the nice slice description they then freeze at $r = r_c$, though evolution would continue in a “natural” slicing[13] approaching $r = 0$. Finally, the evolution produces paired quanta from $r = R$. When their wavelength on the slice is $\ll \omega^{-1}$, the pair is nearly indistinguishable from vacuum, *e.g.* by gravitational scattering[12], but when their wavelength reaches its asymptotic value $\sim \omega^{-1}$, the quanta separate from each other and the horizon and travel into/out of the BH.

Excited quanta typically have $\omega \sim T \sim 1/R$ and are emitted every time $\delta t \sim R$. Thus, in a simple model (see *e.g.* [7]) we consider one species of particle and ignore spin, set $\epsilon = 1/R$, and only keep the states with $k = 1$, so $\omega = 1/R$, and with occupation number zero or one. Then, evolution can be described in time steps of size δt , as follows. First, it suffices to describe evolution of a basis of states for the combined black hole and external Hilbert spaces, $\hat{\mathcal{H}} \otimes \mathcal{H}$. Such a basis can be written in terms of states $|\hat{a}\rangle|a\rangle$, where for example the states $|\hat{a}\rangle$ and $|a\rangle$ are bases for $\hat{\mathcal{H}}$ and \mathcal{H} . The general combined state (and thus general initial state) takes the form $\sum_{\hat{a}a} c_{\hat{a}a} |\hat{a}\rangle|a\rangle$. We describe $|\hat{a}\rangle$ and $|a\rangle$ in terms of strings of qubits, and model evolution for δt as

$$|\hat{a}\rangle|a\rangle \rightarrow \hat{U}|\hat{a}\rangle \otimes \frac{1}{\sqrt{2}} (|\hat{0}\rangle|0\rangle + |\hat{1}\rangle|1\rangle) \otimes U|a\rangle . \quad (2.6)$$

Here, U and \hat{U} are unitary operators that we may think of as describing evolution given by (2.4) of the external wavepackets away from the black hole, and of the internal wavepackets, respectively. Different slicings/mode descriptions yield different U, \hat{U} , but equivalent

evolution. In addition to this evolution, an entangled pair of quanta is produced from the horizon.

When backreaction is included, M decreases an amount $\sim 1/R$ in each time step, and so shrinks to near zero at a time $T_{evap} \sim RS_{BH}$. The late-time external state, with $\mathcal{O}(S_{BH})$ quanta, is found by tracing over internal states in (2.6)(or in (2.3))

$$\rho = \text{Tr}_{\hat{\mathcal{H}}} (|\psi\rangle\langle\psi|) , \quad (2.7)$$

and is mixed, with entropy $S = -\text{Tr}(\rho \log \rho) \sim S_{BH}$. Barring a remnant scenario (which is argued on other grounds[22,23] to be unphysical) the black hole and internal Hilbert space disappear, leaving this mixed state. This is Hawking’s basic argument[2] for information loss – which ultimately conflicts with energy conservation[24].

3. Models of unitary evolution

The basic conflict between QM/energy conservation, locality, and Lorentz/diffeomorphism invariance that arises in the previous scenario has been called the “information paradox.” It was argued in [15,13] (see also [25,11,12]) that this is not a true paradox, in that we don’t have a sharp derivation of the state (2.3) and density matrix (2.7) in a perturbative framework that takes into account backreaction. However, as discussed in [25,11,12,15,13], it seems evident that some amount of nonlocality with respect to the semiclassical picture is needed to avoid the essential argument for lost information. This section will model a kind of nonlocal evolution apparently needed for unitarity, in the effective Hilbert-space approach.

As one guide, we begin with the expectation that the internal Hilbert space of the BH shrinks, and contains no information when $M \rightarrow 0$. A candidate parameterization for $N(M) = \log(\dim \hat{\mathcal{H}})$ is (1.1), but one could also consider other functions decreasing to zero. This contrasts with the evolution (2.6), in which $N(M)$ increases by one in each step (being trivially incautious in distinguishing bits and nats).

We can model unitary evolution, in which $N(M)$ decreases by one at each step, in different modifications of (2.6). One is, separating off the leftmost qubit and choosing a basis element $|\hat{a}\rangle$ for the remaining internal Hilbert space,

$$|\hat{0}\rangle|\hat{a}\rangle|a\rangle \rightarrow \hat{U}|\hat{a}\rangle \otimes |\hat{0}\rangle|0\rangle \otimes U|a\rangle \quad , \quad |\hat{1}\rangle|\hat{a}\rangle|a\rangle \rightarrow \hat{U}'|\hat{a}\rangle \otimes |\hat{0}\rangle|1\rangle \otimes U|a\rangle \quad , \quad (3.1)$$

which is effectively unitary. Moreover, if the rightmost internal qubit is always in a definite state, here $|\hat{0}\rangle$, it can be forgotten or “erased” with impunity. Thus, internal information is transmitted to the external state, and $N(M)$ decreases by one.²

In (3.1) information is relayed from the leftmost of $N(M)$ qubits. In case of a general unitary transformation \hat{U} in the preceding time step, this has no invariant meaning. One may take different models for \hat{U} (also with infalling matter – see below). One is simply $\hat{U} = 1$, in which case the leftmost qubit from evolution (2.6) would not be occupied until a time $t \sim RS_{BH}$, and so information return only begins at this time. Another is a random unitary[14,17,27]. This can be thought of as a model of fast-scrambling[16,28]. Different models may be distinguished[13] by this retention time T_r , describing how long it takes the infalling quanta to mix with the escaping qubits. Note that, even with rapid mixing, the effect of a given infalling quantum initially has tiny effect on the outgoing state, but later in evolution discussion of [16] can apply. Note also that in the semiclassical/Hawking picture (2.6), $T_r = \infty$, and in this sense the longest possible[14] T_r , $\sim RS_{BH}(M)$, is most conservative.

Eq. (3.1) represents a big departure from (2.6), particularly if present in early evolution of the BH. One may consider alternatives with smaller departures. One is

$$\begin{aligned} |\hat{0}\rangle|\hat{0}\rangle|\hat{a}\rangle|a\rangle &\rightarrow \hat{U}|\hat{a}\rangle \otimes \frac{1}{\sqrt{2}}(|\hat{0}\rangle|0\rangle + |\hat{1}\rangle|1\rangle) \otimes U|a\rangle \quad , \quad |\hat{0}\rangle|\hat{1}\rangle|\hat{a}\rangle|a\rangle \rightarrow \hat{U}|\hat{a}\rangle \otimes |\hat{0}\rangle|1\rangle \otimes U|a\rangle \\ |\hat{1}\rangle|\hat{0}\rangle|\hat{a}\rangle|a\rangle &\rightarrow \hat{U}|\hat{a}\rangle \otimes |\hat{1}\rangle|0\rangle \otimes U|a\rangle \quad , \quad |\hat{1}\rangle|\hat{1}\rangle|\hat{a}\rangle|a\rangle \rightarrow \hat{U}|\hat{a}\rangle \otimes \frac{1}{\sqrt{2}}(|\hat{0}\rangle|0\rangle - |\hat{1}\rangle|1\rangle) \otimes U|a\rangle \quad , \end{aligned} \quad (3.2)$$

which, with $\hat{U} = 1$, does not alter the Hawking state until $\sim T_{evap} \sim RS_{BH}$. Other simple generalizations clearly exist, including using other pairs.³ Yet another alternative (also generalizable) uses other states, *e.g.* $|1'\rangle$, $|1''\rangle$, that have small amplitude for occupancy in the Hawking state (2.3):

$$|\hat{q}_1\hat{q}_2\rangle|\hat{a}\rangle|a\rangle \rightarrow \hat{U}|\hat{a}\rangle \otimes \frac{1}{\sqrt{2}}(|\hat{0}\rangle|0\rangle + |\hat{1}\rangle|1\rangle) \otimes |\hat{0}'\hat{0}''\rangle|q'_1q''_2\rangle \otimes U|a\rangle \quad , \quad (3.3)$$

where $q_1, q_2 = 0$ or 1 , independently.

² Note also a relation with the final-state picture of [26]; if a given internal qubit first transitions to a given canonical state, this qubit may then be unitarily forgotten. See also [18].

³ Note that a version where one internal qubit is imprinted in a pair does not relay a net bit of information [7,29].

The unitary evolution laws (3.1)-(3.3) are clearly nonlocal with respect to the semi-classical geometry of the BH. One may worry that there is a more serious objection, namely that anything but the evolution (2.6) (or (2.3)) produces a state that an infalling observer sees as very singular, giving a large departure from expected BH behavior. (The pairing in (2.3) yields cancellations[12] between the contributions of the quanta, interacting with an infalling observer.)

Here, allowing such modest nonlocality can be an asset. In the given basis, we can think of the state of paired quanta near the horizon as being of the form $\frac{1}{\sqrt{2}}(|\hat{0}\rangle|0\rangle + |\hat{1}\rangle|1\rangle)$ until the time these quanta have wavelength $\sim R$, on which time scale the transitions like in (3.1)-(3.3) occur. So, departures from the Hawking state only arise for quanta with wavelengths of order the horizon size. One may imagine these departures as arising from some new nonperturbative, perhaps collective, effect of the quantum BH. This evolution results in modifications to $\mathcal{O}(1)$ quanta per time R , that an infalling observer sees as having energy $1/R$ – apparently for a large black hole, a very small effect. Such effects seem particularly small if they modify, for example, graviton states – these would be essentially undetectable for an infalling observer.

Modifications to LQFT evolution like those described in [13] and modeled here are apparently the minimal required to extract information from a BH. Note that U in (3.1)-(3.3) is taken to describe evolution as in LQFT. In general evolution inside is only expected to be governed by LQFT for a typical time $\sim R$ it takes a quanta to fall into the strong curvature domain; after that we have few constraints on \hat{U} . One can artificially freeze this evolution in a nice slicing. However, this becomes an extreme construction when extended over long times, and in particular a perturbative LQFT quantization on nice slices has been argued to be problematic[15,25,12]. This motivates the departure from LQFT of the types (3.1)-(3.3) (and generalizations), which depart from LQFT in the \hat{U} -evolution for qubits deep inside the BH and being relayed outside the BH, and for the qubits in a region of size $\sim R$ receiving the information. Note also that spacelike communication in *flat* space can be related to acausal communication, by a boost. But, the BH geometry breaks the boost symmetry, fixing a frame with respect to which the evolution can be causal[13].

Evolution of types (3.1)-(3.3) represent different pictures with the common feature that a net one qubit is relayed from inside to outside at each step in time. Eq. (3.1) is a big change to (2.3), in which quanta are not produced paired with internal Hawking excitations. It is a simple model for evolution if true Hawking radiation is not present, such as could occur in a general massive remnant scenario[8] or special cases of it such as

fuzzballs[30]. Eq. (3.2), with $\hat{U} = 1$, postpones significant modification until $t \sim RS_{BH}$ when the BH has shrunk appreciably. After this time, it departs from the Hawking state, but has some of its features, in particular the same average outward flux. And, (3.3) yields extra outward flux, in addition to Hawking evaporation; again with $\hat{U} = 1$ this may be postponed to $t \sim RS_{BH}$. Straightforward generalizations of (3.1)-(3.3) clearly exist.

4. Refinements and enhancements

Various refinements are possible, to bring our models closer to a complete description. First, one can clearly generalize evolution of types (3.1)-(3.3) to incorporate the many different modes in (2.3). There is corresponding flexibility in which modes (3.1)-(3.3) imprint the information, though we may assume they act on modes in a region of size $\mathcal{O}(R)$ near the BH and only on modes close to their asymptotic wavelengths (*i.e.* not highly blueshifted). One may also choose different evolutions corresponding to different rates of reduction of the size $N(M)$ of the BH Hilbert space, though (1.1) seems natural.

One may also incorporate infalling matter, *e.g.* in modes with wavelength $\ll R$. The LQFT description of this is a straightforward extension of the discussion in sec. 2, and we might expect its evolution by \hat{U} until it reaches strong curvature to be approximately that of LQFT.⁴ After this, again there are few constraints on the subsequent \hat{U} evolution on such modes. One attractive possibility is that a particle of wavelength $\ll R$ is “broken up” into a collection of the internal modes described above, before the information is relayed (and unoccupied bits erased) in transformations like (3.1)-(3.3). Note that since the entropy in infalling matter does not exceed $S_{BH}(M)$, we are not in danger of having more information encoded in the BH than can be relayed by these evolution rules.

When faced with a failure of classical mechanics in the atom, Bohr introduced a simple phenomenological model with new rules to capture the correct physical behavior; this led to the development of the profound formalism of quantum mechanics. If the world is quantum mechanical and yet locality holds to a good approximation, we can likewise consider simple models of unitary black hole evolution. Due to the apparent failure of LQFT to describe this evolution, such a model no longer resides within that framework – new principles are needed. These models may clarify constraints on unitary evolution, and may be, as with Bohr’s atom and quantum mechanics, guides to a deeper understanding of the more basic and complete non-local mechanics describing quantum gravity.

⁴ One might also choose slicings staying outside the BH to give holographic[9,10] \hat{U} .

Acknowledgments

I thank S. Mathur, D. Page, and Y. Shi for useful conversations. This work was supported in part by the Department of Energy under Contract DE-FG02-91ER40618 and by grant FQXi-RFP3-1008 from the Foundational Questions Institute (FQXi)/Silicon Valley Community Foundation.

References

- [1] S. W. Hawking, “Particle Creation By Black Holes,” *Commun. Math. Phys.* **43**, 199 (1975) [Erratum-ibid. **46**, 206 (1976)].
- [2] S. W. Hawking, “Breakdown Of Predictability In Gravitational Collapse,” *Phys. Rev. D* **14**, 2460 (1976).
- [3] D. N. Page, “Black hole information,” [hep-th/9305040].
- [4] S. B. Giddings, “Quantum mechanics of black holes,” arXiv:hep-th/9412138
- [5] A. Strominger, “Les Houches lectures on black holes,” arXiv:hep-th/9501071.
- [6] S. B. Giddings, “The black hole information paradox,” arXiv:hep-th/9508151.
- [7] S. D. Mathur, “The Information paradox: A pedagogical introduction,” *Class. Quant. Grav.* **26**, 224001 (2009). [arXiv:0909.1038 [hep-th]]; “What the information paradox is *not*,” [arXiv:1108.0302 [hep-th]].
- [8] S. B. Giddings, “Black holes and massive remnants,” *Phys. Rev. D* **46**, 1347-1352 (1992). [hep-th/9203059].
- [9] G. ’t Hooft, “Dimensional reduction in quantum gravity,” arXiv:gr-qc/9310026.
- [10] L. Susskind, “The World As A Hologram,” *J. Math. Phys.* **36**, 6377 (1995) [arXiv:hep-th/9409089].
- [11] S. B. Giddings, “Locality in quantum gravity and string theory,” *Phys. Rev. D* **74**, 106006 (2006). [hep-th/0604072].
- [12] S. B. Giddings, “Black hole information, unitarity, and nonlocality,” *Phys. Rev. D* **74**, 106005 (2006). [hep-th/0605196].
- [13] S. B. Giddings, “Nonlocality versus complementarity: A Conservative approach to the information problem,” *Class. Quant. Grav.* **28**, 025002 (2011). [arXiv:0911.3395 [hep-th]].
- [14] D. N. Page, “Information in black hole radiation,” *Phys. Rev. Lett.* **71**, 3743 (1993) [arXiv:hep-th/9306083].
- [15] S. B. Giddings, “Quantization in black hole backgrounds,” *Phys. Rev. D* **76**, 064027 (2007) [arXiv:hep-th/0703116].
- [16] P. Hayden, J. Preskill, “Black holes as mirrors: Quantum information in random subsystems,” *JHEP* **0709**, 120 (2007). [arXiv:0708.4025 [hep-th]].
- [17] S. L. Braunstein and K. Zyczkowski, “Entangled black holes as ciphers of hidden information,” [arXiv:0907.1190 [quant-ph]]; S. L. Braunstein, M. K. Patra, “Black hole evaporation rates without spacetime,” *Phys. Rev. Lett.* **107**, 071302 (2011). [arXiv:1102.2326 [quant-ph]].
- [18] B. Czech, K. Larjo, M. Rozali, “Black Holes as Rubik’s Cubes,” [arXiv:1106.5229 [hep-th]].
- [19] J. D. Bekenstein, “Black holes and entropy,” *Phys. Rev. D* **7**, 2333-2346 (1973).

- [20] S. B. Giddings and W. M. Nelson, “Quantum emission from two-dimensional black holes,” *Phys. Rev. D* **46**, 2486 (1992) [arXiv:hep-th/9204072].
- [21] D. A. Lowe, J. Polchinski, L. Susskind, L. Thorlacius and J. Uglum, “Black hole complementarity versus locality,” *Phys. Rev. D* **52**, 6997 (1995) [arXiv:hep-th/9506138].
- [22] S. B. Giddings, “Why aren’t black holes infinitely produced?,” *Phys. Rev.* **D51**, 6860-6869 (1995). [hep-th/9412159].
- [23] L. Susskind, “Trouble for remnants,” [hep-th/9501106].
- [24] T. Banks, L. Susskind and M. E. Peskin, “Difficulties For The Evolution Of Pure States Into Mixed States,” *Nucl. Phys. B* **244**, 125 (1984).
- [25] S. B. Giddings, M. Lippert, “The Information paradox and the locality bound,” *Phys. Rev.* **D69**, 124019 (2004). [hep-th/0402073].
- [26] G. T. Horowitz, J. M. Maldacena, “The black hole final state,” *JHEP* **0402**, 008 (2004). [arXiv:hep-th/0310281 [hep-th]].
- [27] T. Banks, W. Fischler, “Space-like Singularities and Thermalization,” [hep-th/0606260].
- [28] Y. Sekino, L. Susskind, “Fast Scramblers,” *JHEP* **0810**, 065 (2008). [arXiv:0808.2096 [hep-th]].
- [29] S. B. Giddings and Y. Shi, work in progress.
- [30] S. D. Mathur, “Fuzzballs and the information paradox: A Summary and conjectures,” [arXiv:0810.4525 [hep-th]].